The Plant Cell, Vol. 13, 1231–1244, May 2001, www.plantcell.org © 2001 American Society of Plant Physiologists

The Maize *Golden2* Gene Defines a Novel Class of Transcriptional Regulators in Plants

Laura Rossini,[1] Lizzie Cribb,[2] David J. Martin, and Jane A. Langdale[3]

Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, United Kingdom

In the C_4 plant maize, three photosynthetic cell types differentiate: C_4 bundle sheath, C_4 mesophyll, and C_3 mesophyll cells. C_3 mesophyll cells represent the ground state, whereas C_4 bundle sheath and C_4 mesophyll cells are specialized cells that differentiate in response to light-induced positional signals. The *Golden2* (*G2*) gene regulates plastid biogenesis in all photosynthetic cells during the C_3 stages of development. However, G2 function is specifically committed to the differentiation of bundle sheath cell chloroplasts in C_4 leaf blades. In this article, we report the isolation of *G2-like* (*Glk*) genes from maize and rice, providing evidence for a family of *Glk* genes in plants. The expression profiles of the rice *Glk* genes suggest that these genes may act redundantly to promote photosynthetic development in this C_3 species. In maize, *G2* and *ZmGlk1* transcripts accumulate primarily in C_4 bundle sheath and C_4 mesophyll cells, respectively, suggesting a specific role for each gene in C_4 differentiation. We show that G2 and ZmGLK1 both can transactivate reporter gene transcription and dimerize in yeast, which supports the idea that these proteins act as transcriptional regulators of cell-type differentiation processes.

INTRODUCTION

In monocotyledonous grasses such as maize and rice, the leaf is strap shaped and is divided into two distinct regions by the ligule. The proximal portion known as the sheath encloses the stem and provides support to the more distal leaf blade, which is the main photosynthetic organ (Brown, 1958; Freeling, 1992). A series of parallel veins runs longitudinally through the entire leaf, with each vein surrounded by a cylinder of bundle sheath cells that is itself surrounded by mesophyll cells (Brown, 1975). Although this leaf anatomy is shared by all grasses, the function of the bundle sheath and mesophyll cells differs depending on whether plants use the C_4 or C_3 photosynthetic pathway.

In C_3 plants such as rice, CO_2 is fixed by ribulose bisphosphate carboxylase (RuBPCase) in the mesophyll cells, whereas bundle sheath cells act as supporting and conducting tissue. In contrast, C_4 plants such as maize compartmentalize photosynthetic reactions between mesophyll and bundle sheath cells. RuBPCase activity is restricted to bundle sheath cells, whereas mesophyll cells accumulate a set of cell-specific enzymes (phosphoenolpyruvate carboxylase, pyruvate phosphate dikinase, and malate dehydrogenase) that act to fix and then shuttle carbon to the bundle sheath cells (reviewed in Edwards and Walker, 1983). Thus, chloroplasts differentiate for distinct functions in the two cell types and exhibit characteristic dimorphism. Bundle sheath cells have agranal chloroplasts containing large starch granules, whereas mesophyll cells have granal chloroplasts without starch granules (Kirchanski, 1975). C_4 metabolism is effectively mediated by the close physical association of mesophyll and bundle sheath cells. In the maize leaf blade, vein spacing is such that every mesophyll cell is directly adjacent to a bundle sheath cell. However, in leaf-like organs such as husk leaves, the spacing between veins is wider; consequently, some mesophyll cells are not in direct contact with bundle sheath cells. These mesophyll cells accumulate RuBPCase and perform C_3 photosynthesis (and hence are referred to here as C_3 mesophyll cells) (Langdale et al., 1988b).

Phylogenetic studies indicate that the C_4 photosynthetic pathway was derived from the C_3 pathway (Moore, 1982). Within maize, C_3-type photosynthetic differentiation appears to be the ground state, with C_4 specialization requiring additional signals (Langdale et al., 1988b; Langdale and Nelson, 1991). Although the C_4 pathway has been characterized in great detail from a physiological perspective, little is known about the genetic control of the developmental processes associated with it and the evolutionary steps that led to its appearance. To determine the cellular differentiation processes linked to C_4 metabolism, we have identified mutations that

[1] Current address: Università degli Studi di Milano, Dipartimento di Produzione Vegetale, Sezione di Agronomia, Via Celoria 2, 20133 Milan, Italy.

[2] Current address: Genome Biology, Current Science Group, Middlesex House, London W1P 6LB, UK.

[3] To whom correspondence should be addressed. E-mail jane.langdale@plants.ox.ac.uk; fax 44-1865-275147.

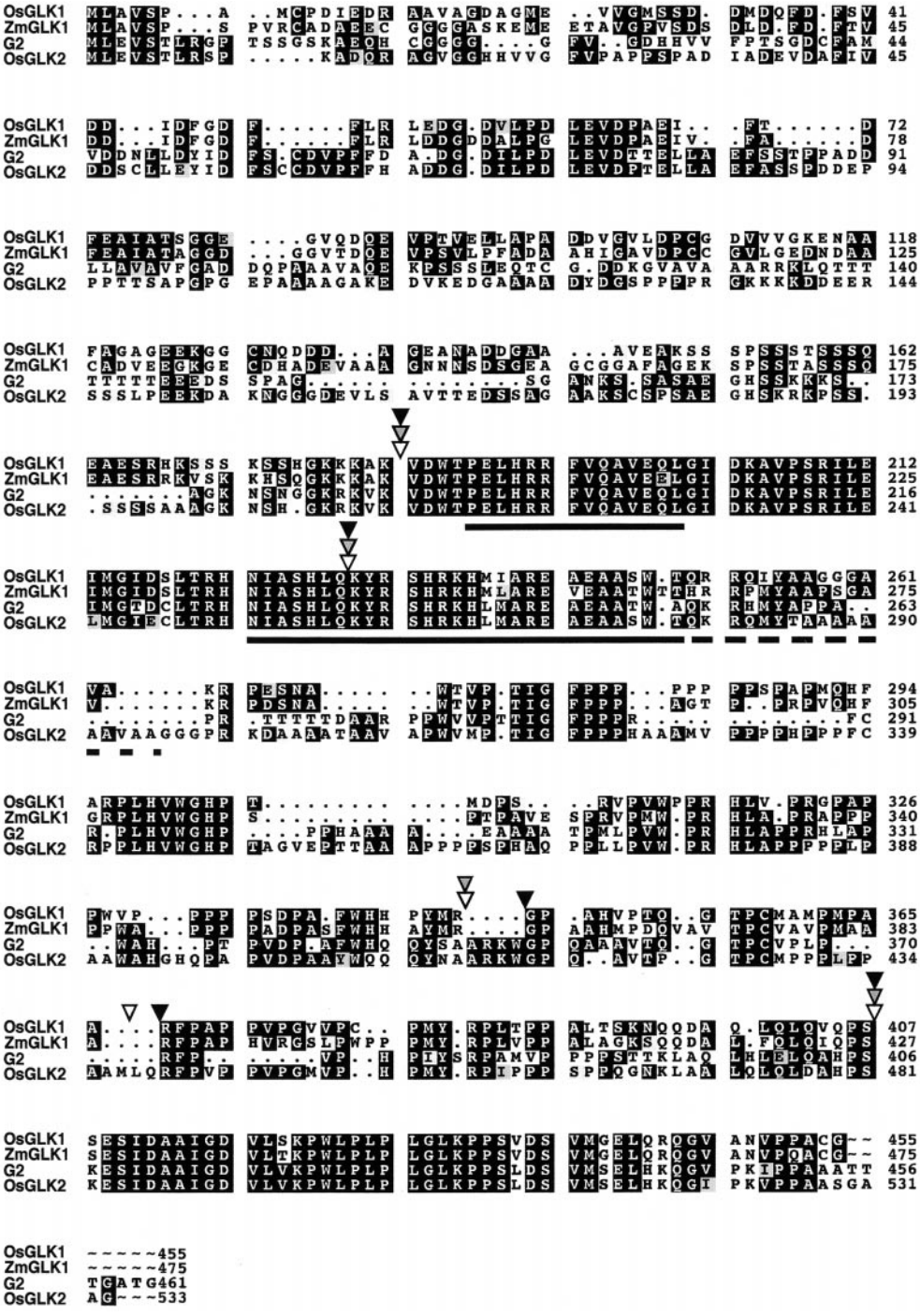

Figure 1. GLK Sequences in Maize and Rice.

Pileup multiple sequence alignment of maize and rice GLK proteins. Amino acid identity is indicated by black boxes, and similarity is indicated by gray boxes. The putative DNA binding domain folds as an HLH: solid horizontal bars indicate the predicted α-helix segments conserved in all four proteins; dashed horizontal bars indicate the maximum extension of the second α-helix, which reaches residue A 295 in OsGLK2, residue A 257 in OsGLK1, residue H 257 in G2, and residue T 263 in ZmGLK1. Triangles mark intron positions in genes: black, *OsGlk1* and *ZmGlk1*; white, *OsGlk2*; gray, *G2*. The GCT box is delimited by the last exon.

disrupt the development of photosynthetic cell types in maize (Langdale and Kidner, 1994; Langdale et al., 1995; Roth et al., 1996). One of these mutations, *golden2* (*g2*), leads to perturbed development of bundle sheath cell chloroplasts in C_4 leaf blades (Langdale and Kidner, 1994). The mutation also disrupts the development of C_3 mesophyll cells, although C_4 mesophyll cells are unaffected. Thus, G2 acts as a general regulator of chloroplast development in C_3-type photosynthetic cells but as a specific regulator of bundle sheath cell chloroplast development in C_4 tissue. The phenotype of *g2* mutant plants further suggests that the loss of *G2* gene function in bundle sheath cells can be compensated for by the action of a second gene (Langdale and Kidner, 1994; L. Cribb and J.A. Langdale, unpublished data).

The *g2* mutant phenotype elicits two testable hypotheses. First, if G2 regulates the development of C_3 mesophyll cells in maize, then *G2*-like genes are likely to be present in C_3 plants. Second, because functional redundancy is widespread in multigene families, a second *G2*-like gene may be present in maize that is able to compensate for the lack of G2 function. The second gene should be expressed primarily in C_4 mesophyll cells after light induction of the C_3-to-C_4 switch. On the basis of the presence of a TEA-related helix-loop-helix (HLH) domain and a functional nuclear localization signal (NLS), G2 was hypothesized to be a transcription factor capable of binding DNA and of forming dimers with itself and other proteins (Hall et al., 1998). To account for redundancy, the second gene should share at least some of these properties.

Here, we show that *G2-like* (*Glk*) gene sequences are present in the C_3 plant rice. Overlapping expression patterns of *OsGlk1* and *OsGlk2* suggest that both genes are involved in photosynthetic differentiation. A second *Glk* gene has been identified in maize. *ZmGlk1* is expressed predominantly in C_4 mesophyll cells in response to light. To substantiate the hypothesis that G2 and ZmGLK1 function as transcriptional regulators, we show that both proteins can transactivate reporter gene transcription in a yeast GAL4 assay. In addition, G2 and ZmGLK1 can form homodimers and heterodimers in the yeast two-hybrid system. *OsGlk1* and *OsGlk2* are highly related to *ZmGlk1* and *G2*, respectively. Sequence relatedness, expression patterns, and the phenotype of *g2* maize mutants suggest that *Glk1* and *Glk2* genes act redundantly to promote photosynthetic development in C_3 species and that the evolution of C_4 involved specialization of *Glk* gene action.

RESULTS

Glk Genes in Maize and Rice

DNA and RNA gel blotting analyses, using the maize *G2* sequence as a probe, suggested that a small family of *Glk* genes is present in rice and maize. Therefore, we screened genomic and cDNA libraries and isolated three *Glk* genes: a maize cDNA named *ZmGlk1* and two rice *G2*-related genes named *OsGlk1* and *OsGlk2*. The maize *ZmGlk1* gene encodes a predicted protein of 475 amino acids (50.4 kD), the rice *OsGlk1* gene encodes a predicted protein of 455 amino acids (48.5 kD), and *OsGlk2* encodes a predicted protein of 533 amino acids (55.9 kD) (Figure 1). Figure 2 shows the percentage amino acid identity derived from GAP pairwise alignments for the peptide sequences encoded by *G2*, *ZmGlk1*, *OsGlk1*, and *OsGlk2*. Sequence identity ranged from 49.3% between the two maize proteins to 72.3% between ZmGLK1 and OsGLK1.

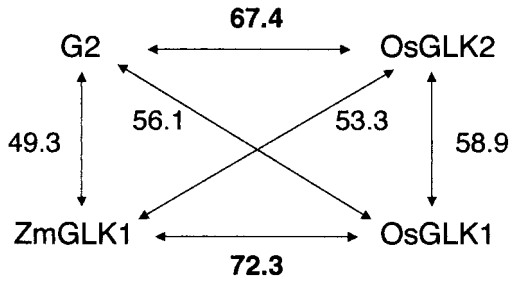

Figure 2. Amino Acid Identities of Maize and Rice GLK Proteins.

Identities were obtained from GAP pairwise alignments between the four grass GLK proteins. The two highest values are shown in boldface.

The multiple sequence alignment shown in Figure 1 demonstrates that sequence conservation is particularly high across two regions. The first region corresponds to the putative DNA binding domain already described for *G2* (Hall et al., 1998). This domain folds as an HLH in PHD computer predictions (Rost, 1996). The first helix, which starts at PELHRR, is invariably 14 amino acids long and is separated from the second helix by a 22–amino acid loop. The second helix starts at the NIASHLQ motif and extends to various lengths in the different GLK proteins. HLH domains bind DNA and mediate dimerization in a number of well-characterized transcriptional regulators (reviewed in Massari and Murre, 2000). The HLH domain in the *Glk* genes shares a low level of sequence similarity with a class of eukaryotic DNA binding domains called TEA, as discussed by Hall et al. (1998; see also references therein). In addition, this domain is highly similar to the B motif described for type B Arabidopsis response regulators (ARRs) (Imamura et al., 1999). ARRs (also called Arabidopsis response regulator-like proteins) are a family of plant proteins related to prokaryotic response regulators of two component systems. The type B subclass has been shown to bind DNA through the B motif (Sakai et al., 1998, 2000; Imamura et al., 1999). A recent publication has assigned both *Glk* and *ARR* genes to a superfamily of transcription factors denoted GARP (G, G2; AR, ARR; P, Psr1; Riechmann et al., 2000; see below). BLAST searches (Altschul et al., 1997) conducted on expressed sequence tag (EST) databases revealed sequences closely related to the GLK HLH domain in a number of other plant species. Table 1 lists

Table 1. GARP Gene Sequences in Plants[a]

Species	Gene	GenBank Accession Numbers
Soybean	1	BE659830, AW119879, BE661803
	2	AW666415, BF008920, BE330035, BF325036, BE330111, BE347643, AW201792
	3	AW202145, BF068309
	3*	BF596099
	4	BF068683, AW596288
	5	AW507631
	5*	AW704492
	6	BE800964
	7	AI522979, AW234794
	7*	AW568114, BE473386
	8	AI495406, AW350460
	9	AI460581, BE801841
	10	AW620443, BF008690
	11	AW100172, AW100166, AW100173
	12	AW318109, BF069630, BF071455
	13	AW509242, BF598100, BE821109
Tomato	1	AI896489, AI782800
	2	BE354404
	3	BE450553, AW217245
	4	AW218302
	5	AW216940, AW441680, AI894498, AI899620, AW032403, AI898971, AI486131, AW223346
	6	AW932398
	7	AW030183
	8	AI488221, AI486663, AI489206, AI487463, AI485303, AI489464
	9	AW222548, AW933286
	10	AW032021
	11	AW222558, AW929913, AW625696, BE462978
	12	AW648066, BE450362
	13	AW030078, AW219675
	14	AI484159
Barley	1	BE559431, BE216742
	2	BF618321
	3	BE519629
	4	BF627268, AL503993
	4*	AL506334
	5	BE519487
	6	BE558795, BE060066
	7	BF617544
	8	BF621611
	9	BF625895
Sorghum	1	AW677774, AW564570
	2	BE366718
	3	BE355874
Cotton	1	AI731854

[a] This list of GARP genes is based on the results of BLAST searches of the nonmouse nonhuman EST database at the National Center for Biotechnology Information (GenBank). ESTs sharing similarity with the GLK HLH domain have been grouped and assembled with other overlapping ESTs to obtain the longest possible contig. Where overlaps were significant but not perfect (due to mismatches or frame shifts between the sequences), a gene number with an asterisk was used to indicate the possible existence of two highly related genes within the same species.

plant GARP genes (excluding those found in Arabidopsis) identified through BLAST searches with GLK protein sequences on the nonmouse nonhuman EST database at NCBI. These genes all have a GLK HLH domain. For each EST containing this domain, subsequent BLAST searches were performed to identify and assemble other overlapping ESTs. Where possible, a list of overlapping ESTs is presented. This analysis confirms that the GARP gene family of transcription factors is widespread in plants, with at least 13 members in soybean (*Glycine max*), 14 in tomato (*Lycopersi-*

con esculentum), nine in barley (*Hordeum vulgare*), three in sorghum (*Sorghum bicolor*), and one in cotton (*Gossypium hirsutum*). In addition, the GLK HLH domain is present in the Ca^{2+}-dependent protein kinase substrate protein1 (CSP1) from *Mesembryanthemum crystallinum* (GenBank accession number AF219972) and in PSR1 (GenBank accession number AF174480), a protein involved in the control of phosphorus metabolism in the green alga *Chlamydomonas reinhardtii* (Wykoff et al., 1999). Notably, no similar sequences exist in the Synechocystis or yeast genomes, and despite extensive database screening, we have not identified this domain in sequences isolated from animal species.

The second conserved region is located in the C-terminal portion of the proteins and is referred to as the GLK/C-terminal box (GCT box). To our knowledge, this domain has not been described previously, and like the HLH, it is not identified by BLAST searches of the yeast and animal databases. Interestingly, only three of the 46 GARP genes identified in the Arabidopsis genome (Riechmann et al., 2000) contain both the HLH and the GCT box (GenBank accession numbers AC007048, AB005239, and AL021889). One of these three (AL021889) encodes an N-terminal receiver domain and thus encodes a "hybrid" GLK and ARR protein. The other two genes resemble those reported here. BLAST database searches (conducted with the same method illustrated above) identified several plant ESTs sharing similarity with the GCT box. At least four genes containing a GCT box were found in soybean. The first is represented by GenBank accession numbers AI495300 and AW350495, the second by accession numbers AW102269 and AW102337, the third by accession number BE348124, and the fourth by accession number AI441555. In the case of the first, significant but not perfect overlap with sequences of soybean gene 1 (as denoted in Table 1) suggests that these overlapping ESTs may represent a member of the *Glk* subfamily. In tomato, the sequence represented by accession number AI782800 contains both HLH and GCT sequences, whereas accession numbers AI484169 and AW223688 represent two independent GCT boxes. GCT boxes are present also in at least one sorghum gene (GenBank accession numbers AW670986 and AW284817) and one barley gene (GenBank accession number BF631280). Notably, the limitations of a survey based on EST assembling (number of EST entries in the database for a given species and modest length and quality of the sequences) do not allow us to accurately identify and assemble all of the *Glk* genes in these species.

In addition to the two conserved regions identified, the G2 sequence contains a recognizable bipartite NLS located at the N terminus of the putative DNA binding domain. Experimental data confirmed that the G2 protein is targeted to the nucleus, at least in a heterologous system (Hall et al., 1998). PSORT computer predictions (Nakai and Kanehisa, 1992) indicate that NLSs also are present in ZmGLK1 and OsGLK2 (Figure 3). The NLS in ZmGLK1 is in a corresponding

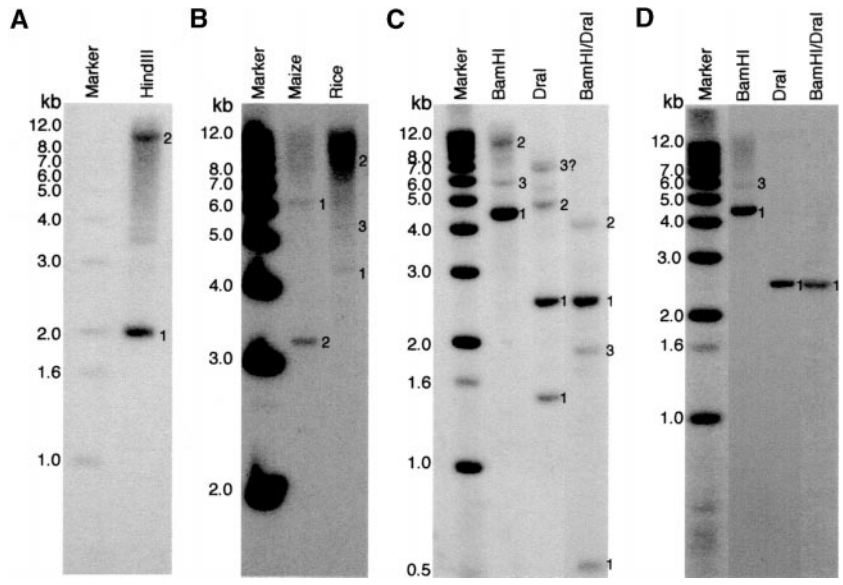

Figure 3. DNA Gel Blot Analysis of *Glk* Genes in Maize and Rice.

(A) HindIII digest of maize genomic DNA hybridized in 0.45 M NaCl at 65°C to the sequence encoding the HLH of G2 (probe 1).
(B) BamHI digests of maize and rice genomic DNA hybridized in 0.75 M NaCl at 65°C to the sequence encoding the HLH of G2 (probe 1).
(C) Digests of rice genomic DNA hybridized in 0.75 M NaCl at 65°C to the sequence encoding the HLH of OsGLK1 (probe 3).
(D) Same filter as in **(C)** hybridized to the 3′ untranslated region (UTR) of *OsGlk1*.
In each case, the gene number is denoted next to the hybridizing fragments.

position and is of a similar type to that identified in G2 (*Xenopus* nucleoplasmin type [Robbins et al., 1991]). In OsGLK2, several overlapping NLSs of the Simian Virus 40 large T antigen type (Kalderon et al., 1984) are found at two sites in exon 1. The first, PPPRGKKKK, is at position 131, and the second, KRKP, at position 188, is located at the N terminus of the putative HLH domain. The PSORT software does not identify an NLS in the OsGLK1 sequence; however, experimental work is required to exclude the possibility that the protein is targeted to the nucleus.

Orthology of Rice and Maize *Glk* Genes

To determine whether we had isolated orthologous *Glk* genes in maize and rice, we first assessed the gene copy number by DNA gel blotting. Maize and rice genomic DNA were digested with either HindIII (Figure 3A) or BamHI (Figure 3B) and hybridized at moderate stringency (0.75 M NaCl at 65°C) to the HLH region of *G2*. Both HindIII and BamHI digests of maize DNA yielded two fragments that hybridized strongly to the *G2* probe. Each was assigned to either *G2* or *ZmGlk1* on the basis of genomic clone restriction maps and hybridization with gene-specific sequences. In addition to the strongly hybridizing fragments, fainter bands and a general smear were observed (Figure 3A). Thus, other more distantly related *Glk* genes are present in the maize genome. Hybridization of the maize *G2* sequence to BamHI-digested rice DNA revealed three genomic fragments of ~4.5, 5.5, and 10 kb (Figure 3B). On the basis of genomic clone restriction maps and hybridization with gene-specific sequences, the most strongly hybridizing 10-kb fragment was assigned to *OsGlk2* and the 4.5-kb species was assigned to *OsGlk1*. To determine whether the 5.5-kb fragment represented a partially digested fragment of *OsGlk1* or a third gene, we performed further restriction digests and hybridized DNA at high stringency (0.45 M NaCl at 65°C) to the HLH region of *OsGlk1* (Figure 3C). In this case, the 4.5-kb BamHI fragment hybridized most strongly, as would be predicted. In double digests with DraI, the presence of a clearly defined 1.8-kb fragment that could not be assigned to either *OsGlk1* or *OsGlk2* and could not represent a partially digested fragment of *OsGlk1* confirmed the presence of a third gene in the rice IR36 genome (Figure 3C). Hybridization with gene-specific fragments demonstrated that *OsGlk3* was more closely related to *OsGlk1* than to *OsGlk2*. Indeed, a gene-specific fragment from the 3′ untranslated region (UTR) of *OsGlk1* hybridized weakly to the 5.5-kb BamHI fragment of *OsGlk3* (Figure 3D). This suggests that the two genes are very similar in sequence and that they may represent a very recent duplication event. This suggestion is further supported by the fact that very few restriction digests distinguished between *OsGlk1* and *OsGlk3* and that the 3′ UTR of *OsGlk1* hybridizes to two transcripts on RNA gel blots (see below).

On the basis of amino acid sequence identity, *OsGlk2* and *G2* appear to be orthologous genes. *ZmGlk1* and *OsGlk1* are also highly related and potentially orthologous, although further investigation is necessary to exclude the possibility that *OsGlk3* is the true ortholog of *ZmGlk1*. To investigate these relationships further, we determined the genomic structure of each *Glk* gene (Figure 4). For *ZmGlk1,* intron positions and sizes were deduced by polymerase chain reaction (PCR), whereas gene structures of *OsGlk1* and *OsGlk2* were determined by comparing genomic and cDNA sequences.

Although introns vary in size from 0.088 to 2.1 kb (data not shown), intron positions appear to be largely conserved. In particular, the first two introns and the last intron are found in exactly the same position in all four genes (Figure 4). Notably, these three introns delimit the two most highly conserved regions corresponding to both the majority of the HLH (exon 2) and the GCT box (last exon). The position of intron 3 just upstream of the AARKW motif at the beginning of exon 4 is precisely conserved between *G2* and *OsGlk2* (Figure 1). This motif is missing from both *OsGlk1* and *ZmGlk1*, in which intron 3 is shifted downstream. The characteristic position of intron 3 and the presence or absence of the AARKW motif support the common ancestry of *G2* and *OsGlk2* and of *ZmGlk1* and *OsGlk1*. However, *ZmGlk1*, *OsGlk1*, and *OsGlk2* all contain an additional intron compared with *G2*. Comparison of *G2* and *Glk* gene structures reveals that exon 4 of *G2* spans exons 4 and 5 of the other *Glk* genes, suggesting that the ancestral *Glk* gene contained the extra intron and that *G2* has lost it. Again, the position of this additional intron is precisely conserved between *ZmGlk1* and *OsGlk1*, confirming that these two genes are more closely related to each other than to *G2* or *OsGlk2*.

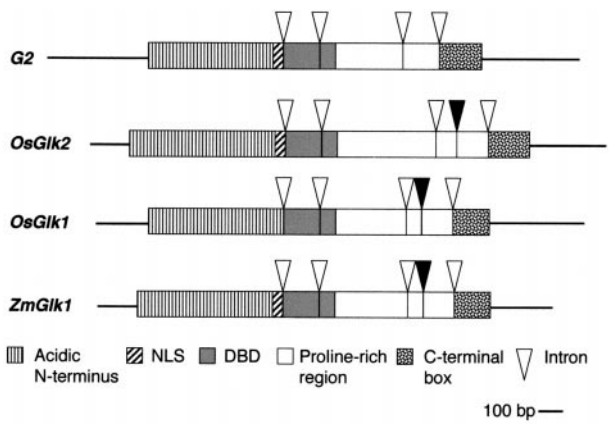

Figure 4. Scheme of *Glk* Gene Structures.

Horizontal lines represent UTRs. Boxes represent different domains of the coding region as indicated. White triangles designate the position of introns present in all four genes, and black triangles designate the position of the intron that is not found in *G2*. DBD, putative DNA binding domain.

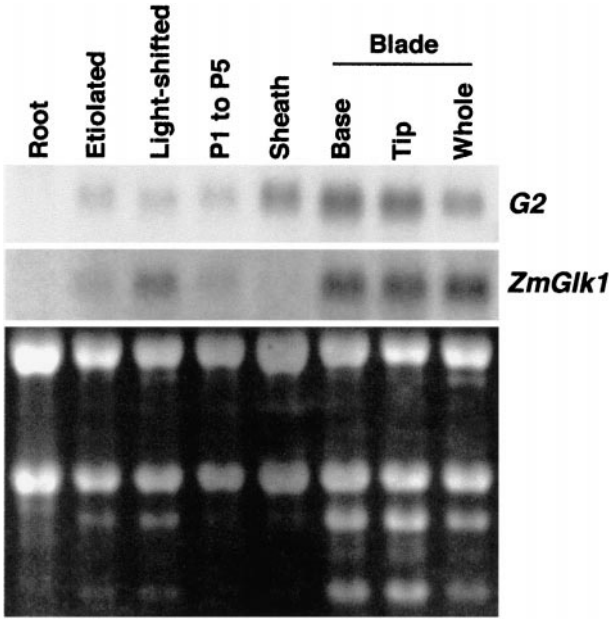

Figure 5. *G2* and *ZmGlk1* Transcript Accumulation in Maize.

RNA gel blot analysis of *G2* and *ZmGlk1* transcript accumulation in maize tissues. *G2* transcript size is 2.2 kb and *ZmGlk1* transcript size is 2.1 kb. The bottom panel shows the ethidium bromide–stained gel to demonstrate RNA loading levels. Blots were hybridized to probe 5 (*G2*) and probe 6 (*ZmGlk1*). P1 to P5, plastochrons 1 to 5.

To further analyze the evolutionary relationships between the maize and rice *Glk* genes, we determined their chromosomal locations by restriction fragment length polymorphism linkage mapping. *G2* maps to the most distal position on the short arm of maize chromosome 3, 1.4 centimorgans (cM) from bnl (tas4L) (Hall et al., 1998). Interestingly, *OsGlk2* maps on the short arm of rice chromosome 1 between C955 (2.2 cM) and S11941A (0.3 cM), a region sharing extensive colinearity with the short arm of maize chromosome 3 (Wilson et al., 1999). The *ZmGlk1* locus is located on the short arm of maize chromosome 9, 1 cM from koln (hox2), and *OsGlk1* is a centromeric marker on rice chromosome 6, which is tightly linked with R688, between R538 (0.8 cM) and P56 (0.2 cM). Notably, maize chromosome 9 and rice chromosome 6 share syntenous segments (Wilson et al., 1999).

Expression Patterns of *Glk* Transcripts

The phenotype of *g2* maize mutants indicates that *G2* facilitates photosynthetic development in both C_3 (sheath) and C_4 (blade) tissue. Consistent with this suggestion, *G2* transcripts accumulate in both sheath and blade tissue and are present in etiolated seedlings that exhibit a C_3 pattern of development but are nonphotosynthetic (Figure 5). *G2* transcripts did not accumulate in roots but accumulated in all photosynthetic tissues examined (Figure 5). Importantly, in mature leaf blades, *G2* transcripts accumulate mainly in C_4 bundle sheath cells, consistent with the specialized role of G2 in regulating bundle sheath chloroplast differentiation (Figures 6A and 6B).

The pattern of *ZmGlk1* transcript accumulation also is consistent with a role in C_4 photosynthetic development. As with *G2*, *ZmGlk1* is not expressed in roots and is expressed throughout leaf development (Figure 5). However, *ZmGlk1* transcripts are barely detectable in leaf sheath tissue. Furthermore, in contrast to *G2*, *ZmGlk1* transcript accumulation is light induced, suggesting a correlation with the switch from C_3 to C_4 photosynthetic development (Figure 5) (Langdale et al., 1988b). In mature leaf blades, the pattern of *ZmGlk1* gene expression is complementary to that of *G2* in that transcripts preferentially accumulate in C_4 mesophyll cells (Figures 6A and 6B).

The rice genes *OsGlk1* and *OsGlk2* also exhibit expression patterns that are consistent with an involvement in photosynthetic development. *OsGlk2* produces a single transcript of 2.3 kb, whereas the *OsGlk1* 3′ UTR sequence hybridizes to two transcripts, a predominant 2-kb mRNA and a larger, very faintly hybridizing 3.3-kb mRNA**.** The 3.3-kb species may represent a partially spliced *OsGlk1* transcript or may represent hybridization to *OsGlk3*, because the probe used was that which hybridizes to *OsGlk3* genomic fragments (Figure 3D). Notably, the 3.3- and 2-kb transcripts exhibited identical accumulation profiles. Transcripts of both genes were absent from roots but were present throughout leaf development. Transcripts accumulated in dark-grown tissue; however, a notable increase in accumulation levels was observed after exposure to light (Figure 7A). The patterns of *OsGlk1* and *OsGlk2* expression essentially overlap, except that *OsGlk1* was more highly expressed in leaves of young seedlings, whereas *OsGlk2* transcript levels were maintained throughout development (Figures 7A and 7B). These overlapping expression profiles suggest that the roles of *OsGlk1* and *OsGlk2* may be similar.

G2 and ZmGLK1 Transactivate Transcription in a Yeast GAL4 Transactivation Assay

A number of observations have suggested that both G2 and the other GLK proteins function by binding DNA and regulating gene expression at the transcriptional level. To substantiate this hypothesis, we tested the ability of G2 and ZmGLK1 peptides to activate transcription in a GAL4 transactivation assay. We constructed chimeras between the GAL4 DNA binding domain (GAL4DB) and the G2 or ZmGLK1 peptides. We then expressed these protein fusions in a yeast strain carrying the *lacZ* reporter gene under the control of a promoter containing GAL4 binding sites. The results showed that full-length sequences of both G2 and ZmGLK1 could activate the transcription of *lacZ* in the yeast system

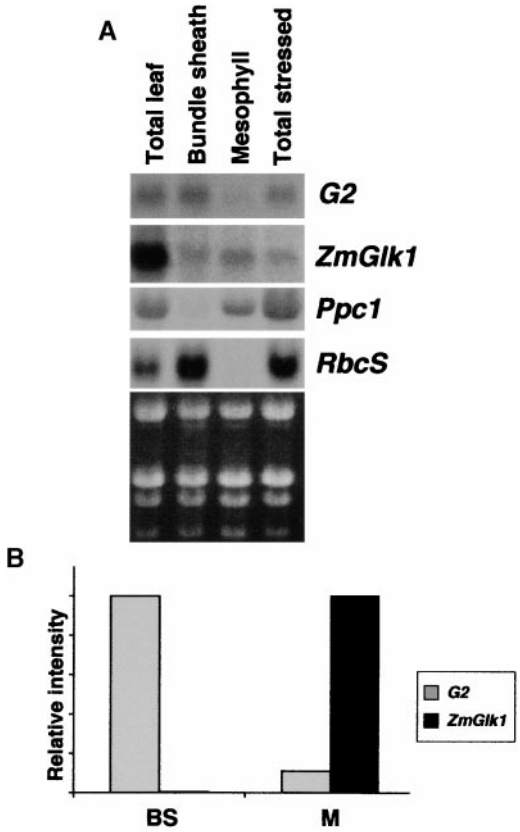

Figure 6. Transcript Accumulation in Maize Bundle Sheath and Mesophyll Cells.

(A) Gel blot analysis of *G2* and *ZmGlk1* transcripts in RNA isolated from purified bundle sheath and mesophyll cells. Blots were hybridized to probe 5 (*G2*) and probe 6 (*ZmGlk1*). Bundle sheath cells were isolated by rapid mechanical disruption. Mesophyll cells were isolated after a 3-hr enzymatic digestion. To monitor transcript degradation during the isolation process, total leaves were incubated as for the mesophyll preparation but in the absence of enzyme. RNA was isolated from these "total stressed'' leaves after 3 hr of incubation. Thus, transcript levels detected in bundle sheath cell samples should be compared with those in "total leaf" samples, whereas levels in mesophyll cell samples should be compared with those in total stressed samples. The purity of cell preparations was assessed by hybridization to mesophyll cell–specific *Ppc1* and bundle sheath cell–specific *RbcS* transcripts. At bottom, the ethidium bromide–stained gel demonstrates RNA loading levels.

(B) Normalization of hybridization signals in bundle sheath (BS) and mesophyll (M) cells after densitometric analysis of autoradiographs. *ZmGlk1* transcripts were degraded during the 3-hr mesophyll isolation procedure, as determined by comparing the hybridization signal in total leaf and total stressed lanes. Thus, hybridization signals for mesophyll cells were normalized to signals in total stressed samples. During the 10-min bundle sheath cell isolation procedure, neither *G2* nor *ZmGlk1* transcripts were degraded, and thus signals in bundle sheath cells were normalized to signals in total leaf samples. Signal density in bundle sheath cell samples is therefore expressed relative to that seen in total leaf samples after adjustment for RNA loading levels. Signal density in mesophyll cell samples is expressed relative to that seen in total stressed samples after adjustment for RNA loading levels.

(Figure 8). In both G2 and ZmGLK1, the N-terminal portion alone, spanning the acidic domain (encoded by exon 1) and the HLH (encoded by exon 2 and part of exon 3), also could activate transcription. In contrast, the C-terminal segment spanning the proline-rich region (downstream of the HLH) and the GCT box (encoded by the last exon) could not function in this assay (Figure 8).

G2 and ZmGLK1 Interact in the Yeast Two-Hybrid System

Because G2 and ZmGLK1 activate transcription in a heterologous system, it is likely that both proteins act as transcriptional regulators. The GAL4 assays indicate that transactivation ability is located in the N-terminal portion of the proteins, as is the putative DNA binding domain. What then is the function of the highly conserved GCT box? One possibility is that the GCT box is required for the homodimerization or heterodimerization of GLK proteins. We tested this idea using the yeast two-hybrid system. Chimeras were constructed between the GAL4 transactivation domain and the G2 or the ZmGLK1 peptides. The ability of constructs to mediate the activation of transcription via interaction with the C-terminal fusions GAL4DB–G2/C or GAL4DB–ZmGLK1/C was then assayed. As reported in Figure 9, both G2 and ZmGLK1 were able to homodimerize. The G2 full-length peptide interacted with the G2 C-terminal peptide, but the G2 C-terminal peptide alone was not sufficient to facilitate this interaction. Evidently, therefore, dimerization is mediated by interactions between different domains of the protein. Similar results were obtained with ZmGLK1. Furthermore, G2 and ZmGLK1 were shown to heterodimerize (Figure 9), but again, the C-terminal peptides were not sufficient to facilitate the interaction. Because the C-terminal region can interact with the full-length protein but not with the N- or C-terminal region alone (despite the fact that the two represent the whole protein), it is likely that either the interacting region spans the two domains or the secondary structure of the truncated proteins precludes dimerization.

DISCUSSION

The *G2* gene was first identified after the isolation of a pale green maize mutant that exhibited perturbed development of photosynthetic cell types (Langdale and Kidner, 1994). Characterization of the *g2* mutant phenotype demonstrated that the gene plays a critical role in the development of chlo-

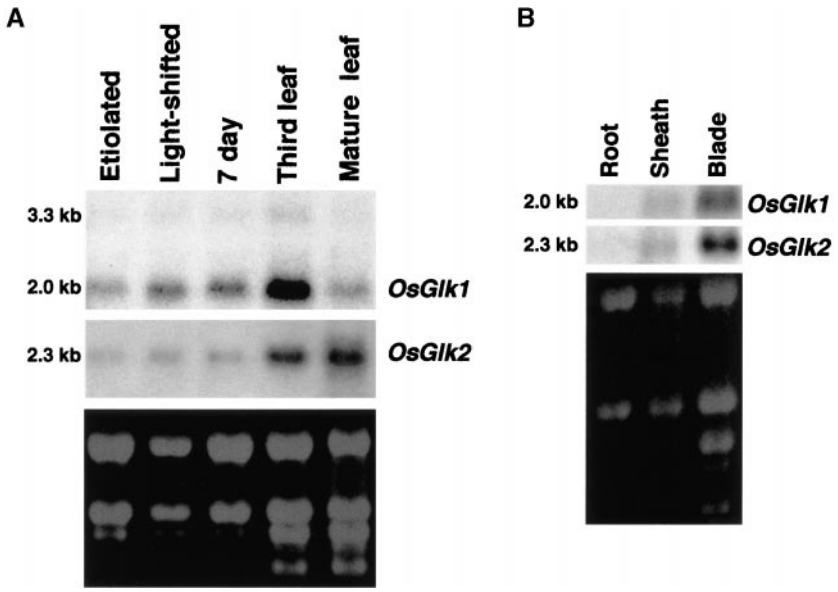

Figure 7. *OsGlk1* and *OsGlk2* Transcript Accumulation in Rice.

(A) RNA gel blot analysis of *OsGlk1* and *OsGlk2* transcript accumulation patterns in rice tissues. Blots were hybridized to probe 2 (*OsGlk1*) and probe 7 (*OsGlk2*). *OsGlk1* 3′ UTR sequence hybridized to two transcripts, the predominant one being 2.0 kb. At bottom, the ethidium bromide–stained gel demonstrates RNA loading levels.

(B) RNA gel blot analysis of *OsGlk1* and *OsGlk2* transcripts in roots and leaves. Only the 2.0-kb *OsGlk1* transcript is shown. At bottom, the ethidium bromide–stained gel demonstrates RNA loading levels.

roplasts in C_3 mesophyll and C_4 bundle sheath cells (Langdale and Kidner, 1994). Sequence analysis suggested that G2 acts as a transcription factor (Hall et al., 1998), and in this study we show that reporter gene transcription is activated in a heterologous system by the N-terminal acidic region of the G2 protein. This observation is consistent with previous reports of acidic transactivation domains in plant transcription factors (reviewed in Schwechheimer et al., 1998).

Many transcription factors are encoded by members of multigene families, and we show here that *Glk* genes are present in more than one copy in both rice and maize. We report the isolation of two genes in each species; however, a third related gene appears to be present in rice, and other more distantly related copies exist in maize, which is an amphidiploid (Gaut and Doebley, 1997). On the basis of comparative sequence analysis, we report here two regions of the GLK proteins that are highly conserved (Figure 1). The first is an HLH region that we assume constitutes the DNA binding domain. This region is shared by many other sequences in the databases, as discussed in Results. However, many of these other sequences do not display the second conserved region, the GCT box, which in the case of all four genes reported here is encoded by a single exon. Because the GCT box plays at least some role in dimerization (Figure 9), we propose that the presence of both the HLH and the GCT box should be the defining feature of *Glk* gene family members. Using this definition, we can identify at least two family members in rice and maize and three family members in Arabidopsis (D. Fitter and J.A. Langdale, unpublished data). Evidence reported here suggests that *OsGlk1* and *OsGlk2* in rice are the orthologs of *ZmGlk1* and *G2* in maize, respectively. Thus, the duplication event that led to specific *Glk1* and *Glk2* genes occurred before the divergence of rice and maize. This conclusion is based on amino acid identity scores between the predicted gene products (Figure 2), intron positions within the genes (in particular the position of intron 3) (Figure 4), and chromosomal map positions. However, on the basis of genomic DNA hybridization profiles to HLH domain probes, a third *Glk* gene appears be present in rice (Figure 3). Hybridization to the 3′ UTR of *OsGlk1* (Figure 3D), similarity of the restriction maps for *OsGlk1* and *OsGlk3*, and mRNA expression patterns detected with the *OsGlk1* 3′ UTR probe (Figure 7) support the idea that *OsGlk3* is more closely related to *OsGlk1* than to *OsGlk2*. One interpretation of this observation is that *OsGlk3* and *OsGlk1* are the result of a recent duplication event that possibly occurred after the divergence of maize and rice. To test this further, we would need to isolate the *OsGlk3* gene.

The four *Glk* genes reported here are expressed exclusively in tissues that differentiate chloroplasts or etioplasts. Thus, given the phenotype of *g2* mutants, it is tempting to

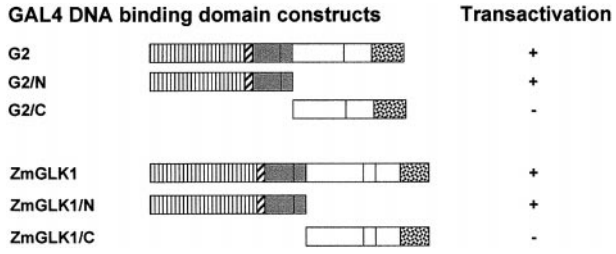

Figure 8. G2 and ZMGLK1 Transactivation Assays in Yeast.

Constructs used in each assay are shown schematically at left. In each case, the regions indicated were fused to the GAL4DB. See Figure 4 and legend for key to the different domains within the *Glk* genes. Positive results in the transactivation assay are represented by (+), and negative results are represented by (−).

speculate that *Glk* genes play a specific role in photosynthetic development. The absence of *Glk* gene sequences in the genome of the cyanobacterium Synechocystis suggests that the presence of *Glk* genes is associated with the ability to develop chloroplasts rather than with the ability to photosynthesize per se. Further support for this idea is provided by the fact that *Glk* gene sequences have been identified only in green algae and in plants. Can *Glk* gene profiles provide insight into differences between C_3 and C_4 photosynthetic development? Because *Glk* genes are duplicated even in C_3 species, the simplest hypothesis would be that *Glk* genes act redundantly to promote the development of all photosynthetic cell types in C_3 plants and that during the evolution of maize, individual family members were recruited for specialized functions.

Because *G2* regulates the development of C_3 cell types in maize, it is reasonable to speculate that the orthologous *OsGlk2* gene may play a similar role in rice. The *OsGlk2* expression pattern also supports this suggestion. Because the expression profile of *OsGlk1* essentially overlaps that of *OsGlk2* (Figure 7), it is possible to speculate further that the two genes act redundantly to promote photosynthetic development. Functional redundancy is a widespread phenomenon in plant genomes and has been shown for other transcription factor families, particularly in the regulation of developmental processes that are crucial for plant survival (reviewed in Martienssen and Irish, 1999). However, on the basis of computer predictions with the PSORT package, the OsGLK1 protein lacks an NLS. If the prediction software is accurate, then this observation could compromise OsGLK1 function, although heterodimerization of OsGLK1 and OsGLK2 would facilitate nuclear import. Although OsGLK1 and OsGLK2 have not been tested for dimerization potential, maize and Arabidopsis GLK proteins all show dimerization capacity in a heterologous yeast system (Figure 9; D.J. Martin and J.A. Langdale, unpublished data). The disruption of rice *Glk* gene function is essential to confirm the involvement of *OsGlk1* and *OsGlk2* in photosynthetic development; however, the evidence obtained thus far supports such a role. In addition, the isolation of *OsGlk3* is important to complete the picture of *Glk* gene function in rice.

On the basis of the *g2* mutant phenotype and work presented here, we propose that the development of distinct bundle sheath and mesophyll cells in C_4 leaf blades requires the complementary action of G2 and ZmGLK1. The development of specialized C_4 bundle sheath and mesophyll cells is proposed to occur in response to a light-induced signal emanating from the veins (Langdale et al., 1988b; Langdale and Nelson, 1991). In the case of bundle sheath cell development, the *g2* mutant phenotype and *G2* gene expression patterns suggest that the *G2* gene functions in this cell type after induction of C_4. Furthermore, G2 regulates the differentiation of C_3 cell types, which in maize include C_3-like etiolated tissue and C_3 mesophyll cells that are situated more than two cells away from a vein. Possibly, the fundamental role of *Glk2/G2* in both C_3 and C_4 species is to promote the development of chloroplasts that accumulate RuBPCase (i.e., C_3 mesophyll and C_4 bundle sheath chloroplasts). The *g2* mutant phenotype suggests that C_4 mesophyll chloroplast development is regulated independently of C_4 bundle sheath chloroplast development. *ZmGlk1* is expressed primarily in C_4 mesophyll cells, is induced by light, and is absent from C_3 mesophyll cells, consistent with the suggestion of a derived role for ZmGLK1 in C_4 plants. An interesting aspect of the *g2* mutant phenotype is the recovery observed in older tissues, which leads to a partial restoration of wild-type chloroplast morphology in bundle sheath cells (Langdale and Kidner, 1994). This was postulated to result from the compensating action of a second gene. Preliminary RNA gel blot analysis demonstrated that *ZmGlk1* transcript levels are normal in *g2* mutants (J.A. Langdale, unpublished data), and thus *ZmGlk1* is an obvious candidate for this role.

Both transactivation and dimerization assays suggest that GLK1 and GLK2 are capable of performing similar functions. In rice, these proteins may act redundantly in all photosynthetic cell types, although some temporal specificity is suggested by the fact that *OsGlk1* is expressed primarily during the early stages of development. In maize, spatial rather than temporal separation of gene expression is observed, with *ZmGlk1* expressed primarily in C_4 mesophyll cells and *G2* expressed primarily in C_4 bundle sheath cells. G2 and ZmGLK1 can form homodimers and heterodimers in a heterologous yeast system, suggesting that various interactions also may occur in vivo. Given the preferential expression of *ZmGlk1* in mesophyll cells and *G2* in bundle sheath cells, it is likely that the formation of homodimers is normally favored. However, low levels of *G2* transcripts are detected in mesophyll cells, and *ZmGlk1* transcripts are detected in bundle sheath cells. Can the action of *Glk* genes account for the differentiation of chloroplasts in all photosynthetic organisms? Presumably, there was a single ancestral *Glk* gene that was duplicated at least once during the evolution of land plants. Although all *Glk* genes may not have been required in C_3 plants, duplication would have permitted further

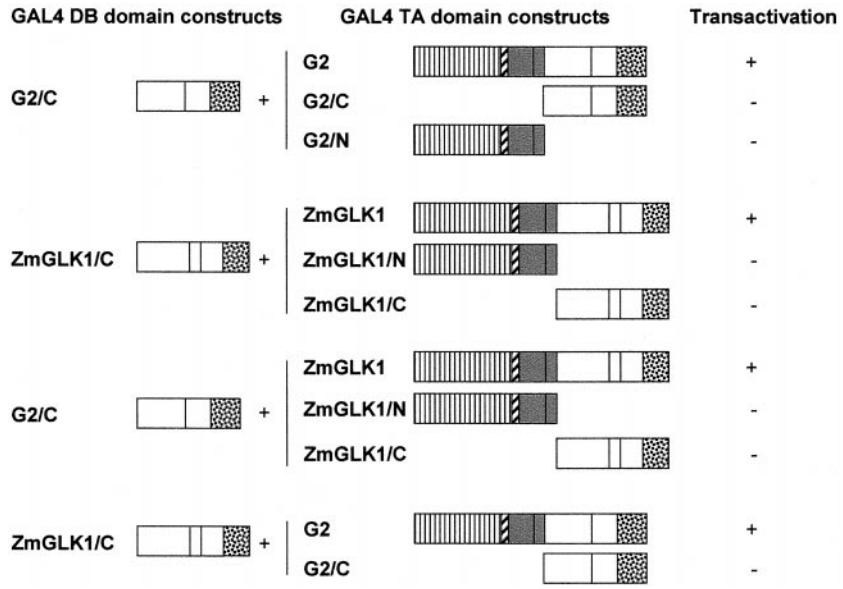

Figure 9. G2 and ZmGLK1 Interaction Assays in Yeast.

GAL4DB constructs are shown schematically at left. These constructs were cotransformed with GAL4 transactivation (TA) domain constructs that included full-length, N-terminal, or C-terminal portions of G2 and ZmGLK1, as represented at center. Positive interactions are indicated by (+), and no interaction is represented by (−).

specialization of the type described here for the C_4 plant maize. This suggestion needs experimental substantiation. Furthermore, loss-of-function mutations need to be identified for *ZmGlk1* and the rice *Glk* genes, and the targets of *Glk* gene action need to be identified to clarify their role in the differentiation process. However, the possibility that *Glk* genes are fundamental regulators of photosynthetic development is an exciting one.

METHODS

Plant Material and Growth Conditions

Maize (*Zea mays*) and rice (*Oryza sativa*) inbred lines B73 and IR36, respectively, were used for all experiments. IR36 rice seedlings were grown in soil in a greenhouse (with an average daytime temperature of 28°C and an average nighttime temperature of 16°C) with a 16-hr-light (supplemented to 500 μmol m^{-2} sec^{-1}) and 8-hr-dark cycle. Etiolated seedlings were grown in the dark for 7 days at 28°C. Light-shifted seedlings were grown in the dark for 7 days and then placed in the greenhouse for 24 hr at the start of the light cycle. Young leaf primordia were harvested 7 days after germination in the greenhouse, when seedlings were ~2.5 cm tall. Third-leaf blades and sheaths were harvested from 4-week-old plants (the fifth leaf was emerging from the main shoot, and some of the plants had one or two tillers). Roots and mature leaves were harvested from 2-month-old plants.

B73 maize seedlings were grown in a growth cabinet, at 28°C constant temperature, with a 16-hr-light (100 μmol m^{-2} sec^{-1}) and 8-hr-dark cycle. Plastochron 1 to 5 leaf primordia were harvested 4 days after planting, when seedlings were ~2 cm tall and all of the leaves were still enclosed in the coleoptile. The young shoots (including the coleoptiles) were excised 2 to 3 mm above the shoot apical meristem. Third leaves and roots from light-grown seedlings were harvested at the stage of emergence of the fourth leaf (~10 days after planting). Leaf sheaths were harvested intact, and leaf blades were divided into basal and distal halves to obtain RNA from base and tip or were harvested intact to obtain whole leaf blade RNA. Etiolated seedlings were grown in vermiculite in complete darkness at 28°C for 7 days. For RNA gel blot analysis, whole seedlings were harvested above the coleoptile ridge. Light-shifted (greening) seedlings were grown in complete darkness, as described above, for 6 days, and then shifted to a growth cabinet at 28°C for a 14-hr-light (100 μmol m^{-2} sec^{-1}), 8-hr-dark, 2-hr-light cycle, before being harvested as described for the corresponding etiolated samples. Purified bundle sheath and mesophyll cells were isolated as described by Hall et al. (1998).

Glk Gene Cloning

A rice genomic library (IR36) was purchased from Clontech (Palo Alto, CA). A rice seedling leaf cDNA library was kindly provided by M. Matsuoka (Nagoya University, Japan). A second rice cDNA library was prepared from 2-month-old greenhouse-grown seedling leaves (IR36 genotype). Total RNA was extracted as described

below, and mRNA was isolated using the PolyATract mRNA isolation system (Promega). The library was constructed in the HybriZAP vector using the HybriZAP two-hybrid cDNA synthesis kit and the HybriZAP two-hybrid cDNA gigapack cloning kit according to the manufacturer's instructions (Stratagene).

Initially, the Matsuoka rice cDNA library was screened at moderate stringency (hybridized in 0.75 M NaCl at 65°C and washed in 0.15 M NaCl at 65°C) with a probe spanning the *G2* cDNA from within the helix-loop-helix (HLH) domain to the end (probe 1, positions 1014 to 2192 in GenBank accession number AF318579). A partial rice cDNA clone was isolated and assigned to the *OsGlk1* gene. Sequence from this clone was used to generate two probes: a gene-specific fragment spanning the 3′ untranslated region (UTR) of *OsGlk1* (probe 2, positions 1550 to 1965 in GenBank accession number AF318581) and a probe spanning the HLH domain of *OsGlk1* (probe 3, positions 763 to 947 in GenBank accession number AF318581). Subsequently, the rice genomic library was screened at moderate stringency (hybridized in 0.75 M NaCl at 65°C and washed in 0.15 M NaCl at 65°C) with both probe 2 and probe 3. Several overlapping clones were isolated and classified into two groups, corresponding to *OsGlk1* and *OsGlk2*, by restriction mapping and hybridization to probe 2. Restriction mapping was subsequently used to identify the most convenient clones for subcloning and sequencing.

Genes were subcloned as follows: pLAR2 is a 4.5-kb BamHI genomic fragment spanning most of the *OsGlk1* locus from a site within the coding region of exon 1 to a site located ~1 kb downstream of the polyadenylation site in the cDNA. pLAR3 is a 1.5-kb DraI genomic fragment spanning the 5′ portion of the *OsGlk1* locus, including exons 1 and 2. pLAR4 is a 4.8-kb DraI genomic fragment spanning the 5′ portion of the *OsGlk2* gene, including exon 1 and exon 2 (sequence from this clone spanning exon 1 is published in GenBank, accession number AF318583). pLAR6 is a 7-kb XhoI genomic fragment spanning most of the *OsGlk2* locus from a site within exon 1 to a site located ~1 kb downstream of the polyadenylation site in the cDNA.

Partial sequence was obtained from these subclones and compared with the *G2* sequence and the *OsGlk1* partial cDNA sequence. Sequence in the first exon was found to be different in *OsGlk1* and *OsGlk2*; therefore, a specific probe for this region of *OsGlk2* was amplified by polymerase chain reaction (PCR) (probe 4, positions 479 to 785 in GenBank accession number AF318583). Our rice cDNA library was then screened with the gene-specific fragments (probe 2 and probe 4) at high stringency (hybridized in 0.45 M NaCl at 65°C and washed in 0.15 M NaCl at 65°C). Fifty independent cDNA clones were isolated for the *OsGlk1* gene. Preliminary characterization by restriction analysis permitted identification of the longest one, pRICE46 (1.9 kb), which was selected for complete sequencing and characterization (GenBank accession number AF318581). Four cDNA clones were isolated for the *OsGlk2* gene, none of which was full length; the longest, pRICEA2 (1.6 kb), was selected for complete sequencing and further characterization (GenBank accession number AF318582). The genomic and cDNA sequences for *OsGlk2* differ in the number of GCC repeats present in the coding region of exon 1, exhibiting three and seven repeats, respectively (translating to three and seven alanine residues, respectively). Sequence for this gene is included in the bacterial artificial chromosome clone OSJNBa0086P08 from rice chromosome 1 (GenBank accession number AP002855). In this sequence, nine GCC repeats are present at this site. A possible interpretation of these discrepancies is that the tandem repeats are prone to expansion/contraction. In this work, the sequence of the *OsGlk2* mRNA was deduced by joining together sequences from the pRICEA2 clone and the pLAR4 genomic subclone, and the number of repeats was derived from the genomic clone on the basis that artifacts are more likely to occur during cDNA synthesis.

Intron positions and size were determined by comparing genomic and cDNA sequences and by PCR. Once approximate positions of introns were established by this strategy, exact positions and junction sequences were obtained by partial sequencing of the genomic subclones.

A maize cDNA library prepared from B73 seedling leaf tissue (a gift from A. Barkan, University of Oregon, Eugene) was screened at moderate stringency (hybridized in 0.75 M NaCl at 60°C and washed in 0.15 M NaCl at 60°C) by using the HLH region of *OsGlk1* as a probe (probe 3). Two clones were isolated, pLAR7 (1.87 kb) and pLAR8 (1.66 kb). Complete sequencing revealed that they represent alternative polyadenylation products of the same gene. pLAR7, being the longer clone, was used in all subsequent experiments (GenBank accession number AF318580). Intron positions and sizes were determined by genomic PCR. PCR products were cloned in pGEMT-easy (Promega) and sequenced using primers to the vector.

Sequence Analysis

The Genetics Computer Group (Madison, WI) GAP and Pileup packages were used for pairwise and multiple alignments, respectively. Various on-line resources were used for retrieval of *Glk*-related sequences. The BLOCKS World Wide Web server at http://www.blocks.fhcrc.org allowed for the detection of conserved blocks among GLK proteins, and Cobbler sequence outputs were used in subsequent BLAST searches. BLAST searches were performed on a range of nucleotide and protein databases at http://www.ncbi.nlm.nih.gov/BLAST/ by using appropriate algorithms with and without low-complexity filters. PHD protein folding predictions were made through the Predict Protein e-mail server at http://www.public.iastate.edu/~pedro/pprotein_query.html. PSORT predictions were made at http://psort.nibb.ac.jp/.

Isolation of DNA and Gel Blot Analysis

DNA was isolated from leaf tissue according to Chen and Dellaporta (1994). DNA gel blot analysis was performed as described by Langdale et al. (1991). Hybridizations were performed in either 0.75 M or 0.45 M NaCl at 65°C, as indicated.

Isolation of RNA and Gel Blot Analysis

RNA was isolated, electrophoresed, blotted, and hybridized as described by Langdale et al. (1988a). Densitometry of autoradiographs and analysis of fluorescence in ethidium bromide–stained gels were performed using the Biorad Fluor-S multiimager and software package. Gene-specific probes used for RNA gel blot analysis were as follows: *G2*, 400-bp 3′ UTR fragment from the LNH10 subclone (Hall et al., 1998) (probe 5, positions 1809 to 2192 in Genbank accession number AF318579); *ZmGlk1*, last 240 bps of the pLAR7 cDNA spanning the 3′ UTR (probe 6, positions 1601 to 1843 in GenBank accession number AF318580); *OsGlk1*, last 415 bps of the pRICE46 cDNA spanning the 3′ UTR (probe 2); *OsGlk2*, 303 bps of the pRICEA2 cDNA starting at the stop codon and spanning the 3′ UTR (probe 7, positions 1291 to 2098 in GenBank accession number AF318582).

Yeast GAL4 Assays

The GAL4 transactivation assays and the yeast two-hybrid assays were performed using the MATCHMAKER two-hybrid system from Clontech. GAL4 DNA binding domain fusions were constructed in the pGBT9 vector as follows: G2 full length, from the ATG (residue M 1) to the stop codon; G2 N terminus, from the ATG (residue M 1) to position 1191 in the cDNA (residue A 260); G2 C terminus, from position 1192 in the cDNA (residue P 261) to the stop codon; ZmGLK1 full length, from the ATG (residue M 1) to the EcoRI site at position 1671 in the 3′ UTR; ZmGLK1 N terminus, from the ATG (residue M 1) to position 954 in the cDNA (residue H 264); ZmGLK1 C terminus, from position 955 in the cDNA (residue R 265) to the EcoRI site at position 1671 in the 3′ UTR.

Restriction sites were introduced by PCR amplification with appropriately designed primers at the 5′ end of each fragment to maintain the reading frame of the GAL4 DNA binding domain. The empty pGBT9 vector was included in the experiments as a negative control.

GAL4 transactivation domain fusions were constructed in the pGAD424 vector by using the same fragments as described for the pGBT9 constructs.

Yeast transformation, selection of transformants, and filter β-galactosidase assays were performed as described in the MATCHMAKER two-hybrid system manual (Clontech).

Homodimerization and heterodimerization assays were conducted by cotransforming the GAL4 fusion constructs into yeast. Interactions were scored using the β-galactosidase assay. As a control, pGAD424 derivatives were transformed alone to exclude the possibility that GLK peptides bind to GAL4 binding sites. In addition, the pGBT9–GLK C-terminal fusions were cotransformed with empty pGAD424 vector to show that GLK proteins do not interact directly with the GAL4 transactivation domain. All these controls gave negative results (no transactivation of the reporter gene).

ACKNOWLEDGMENTS

We thank Benn Burr (Brookhaven National Laboratory, Upton, NY), Masahiro Yano (Rice Genome Project, Japan), and Susan McCouch (Cornell University, Ithaca, NY) for help with mapping experiments. We also thank other members of the group, Tom Brutnell, and Miltos Tsiantis for many useful discussions. L.R. was supported by a European Molecular Biology Organization Long Term Fellowship and by a fellowship from the Fondazione Valeria Vincenzo Landi of the Accademia Nazionale dei Lincei (Italy), and L.C. was supported by a Biotechnology and Biological Science Research Council (BBSRC) postgraduate studentship. Work in the laboratory was funded by the BBSRC and by the Gatsby Charitable Foundation.

Received November 6, 2000; accepted March 19, 2001.

REFERENCES

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. **25,** 3389–3402.

Brown, W.V. (1958). Leaf anatomy in grass systematics. Bot. Gaz. **119,** 170–178.

Brown, W.V. (1975). Variations in anatomy, associations, and origins of Kranz tissue. Am. J. Bot. **62,** 395–402.

Chen, J., and Dellaporta, S.L. (1994). Urea-based plant DNA miniprep. In The Maize Handbook, M. Freeling and V. Walbot, eds (New York: Springer-Verlag), pp. 526–528.

Edwards, G.E., and Walker, D.A. (1983). *C3*, *C4*: Mechanisms and Cellular and Environmental Regulation of Photosynthesis. (Oxford: Blackwell Scientific Publications).

Freeling, M. (1992). A conceptual framework for maize leaf development. Dev. Biol. **153,** 44–58.

Gaut, B.S., and Doebley, J. (1997). DNA sequence evidence for the segmental allotetraploid origin of maize. Proc. Natl. Acad. Sci. USA **94,** 6809–6814.

Hall, L.N., Rossini, L., Cribb, L., and Langdale, J.A. (1998). GOLDEN2: A novel transcriptional regulator of cellular differentiation in the maize leaf. Plant Cell **10,** 925–936.

Imamura, A., Hanaki, N., Nakamura, A., Suzuki, T., Taniguchi, M., Kiba, T., Ueguchi, C., Sugiyama, T., and Mizuno, T. (1999). Compilation and characterization of *Arabidopsis thaliana* response regulators implicated in His-Asp phosphorelay signal transduction. Plant Cell Physiol. **40,** 733–742.

Kalderon, D., Roberts, B.L., Richardson, W.D., and Smith, A.E. (1984). A short amino acid sequence able to specify nuclear location. Cell **39,** 499–509.

Kirchanski, S.J. (1975). The ultrastructural development of the dimorphic plastids of *Zea mays* L. Am. J. Bot. **62,** 695–705.

Langdale, J.A., and Kidner, C.A. (1994). *bundle sheath defective*, a mutation that disrupts cellular differentiation in maize leaves. Development **120,** 673–681.

Langdale, J.A., and Nelson, T. (1991). Spatial regulation of photosynthetic development in C_4 plants. Trends Genet. **7,** 191–196.

Langdale, J.A., Rothermel, B.A., and Nelson, T. (1988a). Cellular patterns of photosynthetic gene expression in developing maize leaves. Genes Dev. **2,** 106–115.

Langdale, J.A., Zelitch, I., Miller, E., and Nelson, T. (1988b). Cell position and light influence C_4 versus C_3 patterns of photosynthetic gene expression in maize. EMBO J. **7,** 3643–3651.

Langdale, J.A., Taylor, W.C., and Nelson, T. (1991). Cell-specific accumulation of maize phosphoenolpyruvate carboxylase is correlated with demethylation at a specific site >3 kb upstream of the gene. Mol. Gen. Genet. **225,** 49–55.

Langdale, J.A., Hall, L.N., and Roth, R. (1995). Control of cellular differentiation in maize leaves. Philos. Trans. R. Soc. Lond. (Biol.) **350,** 53–57.

Martienssen, R., and Irish, V. (1999). Copying out our ABCs: The role of gene redundancy in interpreting genetic hierarchies. Trends Genet. **15,** 435–437.

Massari, M.E., and Murre, C. (2000). Helix-loop-helix proteins: Regulators of transcription in eucaryotic organisms. Mol. Cell Biol. **20,** 429–440.

Moore, P.D. (1982). Evolution of photosynthetic pathways in flowering plants. Nature **295,** 647–648.

Nakai, K., and Kanehisa, M. (1992). A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics **14,** 897–911.

Riechmann, J.L., et al. (2000). *Arabidopsis* transcription factors: Genome-wide comparative analysis among eukaryotes. Science **290,** 2105–2110.

Robbins, J., Dilworth, S.M., Laskey, R.A., and Dingwall, C. (1991). Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: Identification of a class of bipartite nuclear targeting sequence. Cell **64,** 615–623.

Rost, B. (1996). PHD: Predicting one-dimensional protein structure by profile-based neural networks. Methods Enzymol. **266,** 525–539.

Roth, R., Hall, L.N., Brutnell, T.B., and Langdale, J.A. (1996). *bundle sheath defective2*, a mutation that disrupts the coordinated development of bundle sheath and mesophyll cells in maize. Plant Cell **8,** 915–927.

Sakai, H., Aoyama, T., Bono, H., and Oka, A. (1998). Two-component response regulators from *Arabidopsis thaliana* contain a putative DNA-binding motif. Plant Cell Physiol. **39,** 1232–1239.

Sakai, H., Aoyama, T., and Oka, A. (2000). *Arabidopsis* ARR1 and ARR2 response regulators operate as transcriptional activators. Plant J. **24,** 703–711.

Schwechheimer, C., Smith, C., and Bevan, M.W. (1998). The activities of acidic and glutamine-rich transcriptional activation domains in plant cells: Design of modular transcription factors for high-level expression. Plant Mol. Biol. **36,** 195–204.

Wilson, W.A., Harrington, S.E., Woodman, W.L., Lee, M., Sorrells, M.E., and McCouch, S.R. (1999). Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated panicoids. Genetics **153,** 453–473.

Wykoff, D.D., Grossman, A.R., Weeks, D.P., Usuda, H., and Shimogawara, K. (1999). Psr1, a nuclear localized protein that regulates phosphorus metabolism in Chlamydomonas. Proc. Natl. Acad. Sci. USA **96,** 15336–15341.